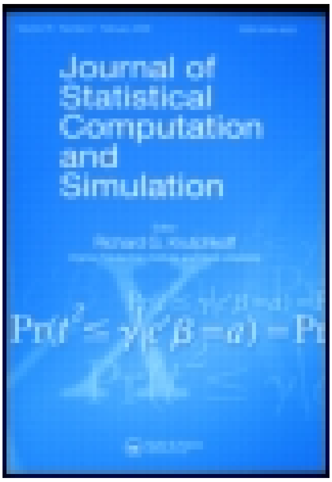


Journal of Statistical Computation and Simulation
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/gscs20

# Improved simultaneous intervals for linear combinations of parameters from generalized linear models

Amy Wagler[a] & Melinda McCann[b]
[a] Mathematical Sciences, The University of Texas at El Paso, 500 W. University Ave., El Paso, TX 79968, USA
[b] Statistics, Oklahoma State University, Stillwater, OK 74074, USA
Published online: 01 Jul 2014.

*Journal of Statistical Computation and Simulation*, 2014
http://dx.doi.org/10.1080/00949655.2014.933339

# Improved simultaneous intervals for linear combinations of parameters from generalized linear models

Amy Wagler[a]* and Melinda McCann[b]

[a]*Mathematical Sciences, The University of Texas at El Paso, 500 W. University Ave., El Paso, TX 79968, USA;* [b]*Statistics, Oklahoma State University, Stillwater, OK 74074, USA*

(*Received 28 January 2014; accepted 7 June 2014*)

When employing generalized linear models, interest often focuses on estimation of odds ratios or relative risks. Additionally, researchers often make overall conclusions, requiring accurate estimation of a set of these quantities. Consequently, simultaneous estimation is warranted. Current simultaneous estimation methods only perform well in this setting when there are a very small number of comparisons and/or the sample size is relatively large. Additionally, the estimated quantities can have significant bias especially at small sample sizes. The proposed bounds: (1) perform well for a small or large number of comparisons, (2) exhibit improved performance over current methods for small to moderate sample sizes, (3) provide bias adjustment not reliant on asymptotics, and (4) avoid the infinite parameter estimates that can occur with maximum-likelihood estimators. Simulations demonstrate that the proposed bounds achieve the desired level of confidence at smaller sample sizes than previous methods.

**Keywords:** generalized linear models; simultaneous inference; multiple comparisons; type I error control

*AMS Subject Classification*: 62J12; 62J15

## 1. Introduction

Modern epidemiological and medical research routinely employs generalized linear models (GLMs) for quantifying relationships between the incidence of disease and particular risk factors. These models can be helpful in understanding what behaviours or traits can influence the incidence of a particular characteristic.

For illustration, data from the 2009 National Health Interview Survey (NHIS) [1] provide information about childhood asthma and other health conditions affecting US youth. For the asthma data, a GLM could be utilized to assess the impact of region and hayfever allergy status on the incidence of asthma-related emergency room (ER) visits for U.S. children. For this analysis, the response variable is binary, recording whether each child had visited the ER due to an asthma attack in the past 12 months and is predicted using reference-coded explanatory variables indicating the region of the USA where the child resides ($x_1$ is an indicator for the Midwest, $x_2$ for the South, $x_3$ for the West, leaving the reference level as the Northeast) and diagnostic status for hayfever allergies ($x_4$ is an indicator for a diagnosis of hayfever allergies). Confidence intervals for the odds ratios (OR) comparing subjects with or without hayfever allergies in different regions of the USA could be estimated using confidence intervals for linear combinations of the slope parameters.

*Corresponding author. Email: awagler2@utep.edu

For example, $\beta_1 - \beta_2$ is the log OR for comparing subjects without hayfever allergies in the Midwest to subjects without hayfever allergies in the South if $\beta_1$ and $\beta_2$ are the slopes for $x_1$ and $x_2$, respectively. The number of these comparisons that are of interest simultaneously to the researcher can easily become quite large. For instance, comparing across the regions separately for each allergy status results in $\binom{4}{2} \times 2 = 12$ comparisons. Even larger sets of simultaneous inferences could reasonably be specified for this example.

Requiring simultaneous inference for a large set of comparisons is not unique to this setting. For instance, pairwise comparisons between the levels of the explanatory variables or comparisons with a reference level are reasonable for many different applications. Alternatively, the researchers may wish to: (1) select a subset of the ORs that significantly lower or raise the odds of a particular outcome or (2) order the ORs for all combinations of a set of discrete explanatory variables. If the usual 95% confidence intervals are used for ordering or comparing the ORs in either of the aforementioned manners, the assumed overall error rate is inflated. For instance, in the asthma example, the overall error for the estimated ORs could be as high as 46% as these would require compiling information from multiple intervals to make overall conclusions. Since researchers rarely desire to limit their conclusions to those involving a single inference (i.e. a single comparison of log ORs), a procedure for controlling multiplicity among the set of inferences is necessary and, ideally, should do so for a large discrete set of inferences while maintaining good power and avoiding estimation difficulties. With these goals in mind, this manuscript presents a procedure for calculating simultaneous intervals on the model parameters of GLMs which has the added benefits of performing well at small sample sizes, reducing bias, and avoiding estimation problems common in this setting. In the following, an overview of the GLM is provided along with an illustration of a set of linear combinations of the model parameters that we seek to estimate simultaneously. After estimating a GLM, like the one modelling the occurrence of asthma-related ER visits, it is typically of interest to estimate particular quantities including mean responses, ORs, or relative risks (RRs). Customarily, these are reported via confidence intervals using some prespecified level of significance for each inference. These confidence bounds may be used to estimate a linear combination of the model parameters.

When a discrete set of inferences is the focus (as in the asthma example), one standard approach for a multiplicity correction is Bonferroni's adjustment. This has the advantage of being simple, but is often too conservative and can lack power, particularly when the number of inferences is large and there are correlations among the inference set. The Sidak adjustment [2] provides a slight improvement over the Bonferroni adjustment, but can be a conservative method when the outcomes are correlated. Similarly, the Hunter–Worsley correction is less conservative than the Bonferroni adjustment because it takes into account correlations among the inference set. However, it is still generally conservative.[3,4] There are also related step-up or step-down procedures that may be applied to discrete sets when ordering or ranking is desired. These procedures are useful and improve power; however, our interest in this manuscript is in simultaneous interval estimators that also provide an assessment of the practical significance of the effect. For standard applications of these methods, the parameter estimates are required to be normally distributed. For GLM settings, the estimators are asymptotically normal.

When the Bonferroni or Sidak procedures perform poorly, many rely on a Scheffé adjustment. Scheffé-based adjustments may be applied to a discrete set of comparisons, but really are not ideal for this setting as Scheffé adjustments allow infinitely many comparisons. Thus, if a finite set are planned, this will often be too conservative. Though Piegorsch and Casella [5] and Casella and Strawderman [6] and, recently, Wagler and McCann [7] have made modifications to the usual Scheffé critical value to improve its precision, these are still conservative methods even with the improvements. Consequently, when a moderate to large number of comparisons are desired, the simultaneous confidence regions (SCRs) developed by Sun et al. [8] are often a more powerful alternative. These bands, first developed for linear regression models, were later adapted with an additional bias correction for estimating the mean response of GLMs.[8] The corrections utilize inverse Edgeworth expansions on the model-based Gaussian random field so that asymptotically the intervals achieve simultaneous coverage for the response of a GLM. However, even with this bias correction, the SCRs may perform poorly, particularly in scenarios with small to medium sample sizes or when there are multiple categorical explanatory variables.

Another complicating issue in this setting is that GLMs with multiple categorical predictors are particularly prone to separability or quasi-separability. This occurs when there is a hyperplane in the explanatory variable space that perfectly separates the classes of the predicted variable. When separability occurs, the MLE-based parameter estimates are not reasonable and another estimation procedure must be utilized. Even when the MLE is reasonable, the GLM with categorical predictors is very prone to bias in the parameter estimates, particularly for small sample sizes. Since our objective is to make reasonable and precise simultaneous estimates for linear combinations of the model parameters, we wish to adjust for the bias and avoid infinite estimates in these settings. Consequently, we propose simultaneous bounds that are similar to SCRs but adjust for the bias in the estimator using a non-informative prior to penalize the log likelihood. Since the bias correction does not rely on asymptotics, these bounds should perform better than the bias-corrected SCRs at small to moderate sample sizes.

In order to further investigate the use of simultaneous methods in GLM settings, we consider GLMs with $k > 1$ categorical explanatory variables. This is a frequently utilized model in biomedical applications since it explores the relationships between a condition and multiple risk factors, often by estimating a set of ORs or RRs summarizing these relationships. Thus, our focus is on simultaneous estimation of various linear combinations of the model parameters. We first demonstrate that the SCRs developed by Sun et al. [8] may be applied to GLMs with categorical predictors and investigate the performance of these intervals in this setting.

In summary, this manuscript builds upon previous research by focusing on multiple sets of inference in GLMs where the explanatory variables are categorical, utilizing a penalization in the estimation procedure in order to improve the small sample performance and adapting the continuous domain approach (i.e. tube formulas as used in the SCRs of Sun et al. [8]) for a family of discrete domain comparisons. In the following sections, we (2) review estimation procedures for GLMs, (3) present relevant simultaneous estimation methods for GLMs, (4) propose penalized SCRs, (5) present simulation results comparing various relevant methods to the proposed method, (6) present an application of the proposed method, and (7) provide concluding remarks and recommendations.

## 2. Estimation of the GLM parameters

In a GLM, the independent response $y_i$, $i = 1, \dots, n$, is modelled

$$f(y_i; \eta_i) = \exp\left\{ y_i\eta_i - \frac{b(\eta_i)}{a(\phi_i)} + c(y_i, \phi_i) \right\}, \tag{1}$$

where $a(\cdot)$, $b(\cdot)$, and $c(\cdot)$ are known functions.[9] The linear predictor of a GLM is a transformation of the mean ($\mu_i = E(y_i)$). This is given by $\eta_i = \boldsymbol{x}_i'\boldsymbol{\beta}$, where $\boldsymbol{\beta}$ is a $p \times 1$ vector of the model parameters and $\boldsymbol{x}_i$ is a $1 \times p$ vector containing the covariates for $i = 1, \dots, n$. The link, $g$, relates the mean to the linear predictor via $g(\mu_i) = \boldsymbol{x}_i'\boldsymbol{\beta}$. When the dispersion parameter $\phi$ is known, this is exponential class.

The logistic regression model, the log-linear Poisson model, the probit model, and the complementary log–log model are all GLMs where simultaneous estimation of functions of the model parameters are often of interest. Any of these models can be expressed as in Equation (1). In this and later sections, $\boldsymbol{Y} = (y_1, \ldots, y_n)'$ is the vector of responses and $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)'$ is the full rank matrix of predictor variables. We denote the usual estimator for the model parameters, the maximum-likelihood estimate (MLE), as $\hat{\boldsymbol{\beta}}_m = (\hat{\beta}_{m_1}, \ldots, \hat{\beta}_{m_k})$. When utilizing the MLE, two-sided bounds for a set of linear functions of the model parameters are given by

$$(\boldsymbol{C}\hat{\boldsymbol{\beta}}_m \pm d\hat{\sigma}_m), \tag{2}$$

where $\hat{\boldsymbol{\beta}}_m$ is the MLE, $\boldsymbol{C}$ is a real-valued matrix, and $\hat{\sigma}_m = \sqrt{\boldsymbol{C}'\boldsymbol{V}_m\boldsymbol{C}}$ is the asymptotic variance of $\hat{\boldsymbol{\beta}}_m$ with $I(\hat{\boldsymbol{\beta}}_m)$ the Fisher information matrix for $\hat{\boldsymbol{\beta}}_m$ and $\boldsymbol{V}_m = I^{-1}(\hat{\boldsymbol{\beta}}_m)$. Though not the focus in this manuscript, lower or upper bounds may be obtained via $\boldsymbol{C}\hat{\boldsymbol{\beta}}_m - d_l\hat{\sigma}$ or $\boldsymbol{C}\hat{\boldsymbol{\beta}}_m + d_u\hat{\sigma}$, respectively. The constants, $d$, $d_l$ and $d_u$, are approximate quantiles from a standard normal distribution and are chosen for some prespecified level of coverage.

Utilizing the MLE can, at times, be problematic. For example, at smaller sample sizes and when the explanatory variables are categorical, the data are more likely to be linearly separable. In this case, the MLE can yield infinite parameter estimates. Additionally, even when the parameter estimates are finite, the MLE is known to yield biased estimates for small sample sizes [10] and the error distribution is often not near normal except for very large sample sizes.[8] For these reasons, we also consider the penalized maximum-likelihood estimate (pMLE) and a weakly informative prior estimate [11] as alternatives to the MLE when constructing simultaneous interval estimates for functions of the parameters for GLMs.

### 2.1. *Incorporating Bayesian priors into maximum-likelihood estimation*

Although still operating from a frequentist standpoint, we propose utilizing estimates that originate from a Bayesian perspective, with emphasis placed on correcting common estimation problems for logistic and Poisson models. The type of Bayesian-influenced estimator can vary, but in this section, two are offered that make few or no assumptions about the model parameters. Others have also recommended routine use of Bayesian estimates in logistic regression models. In particular, Gelman et al. [11] and Firth [10] have proposed utilizing priors for the parameters because, due to separability, the MLE is often unstable and yields infinite parameter estimates. Firth [10] was an early advocate for parameter estimation in GLMs derived by considering priors and others have proposed similar kinds of estimates.[11] The pMLE, introduced by Firth, utilizes a non-informative prior and shifts the usual log-likelihood function using this ‘penalty’ (i.e. Jeffrey’s prior). Applying this penalty has the effect of avoiding non-convergent solutions resulting from separable data and also reduces the bias present in MLEs. A similar approach suggested by Gelman et al. [11] also utilizes a Bayesian perspective for parameter estimation in logistic models. The approach advocated by Gelman et al. incorporates additional prior information that may be assumed for most generic GLMs.

Either the Firth or Gelman methods described above may be implemented by utilizing iteratively weighted least squares (IRLS) to compute the parameter estimates. The details of the estimation procedure may be found in [10,11]. Utilizing the notation outlined above, when computing the pMLE (i.e. Firth’s method) the first derivative of the likelihood is adjusted in the following manner. The first derivative of the likelihood is $(\boldsymbol{D}^{\mathbf{T}}\boldsymbol{u})^* = \partial l^*/\partial\boldsymbol{\beta} = \partial l/\partial\boldsymbol{\beta} + \mathbf{i}(\boldsymbol{\beta})\boldsymbol{b}_{\mathbf{1}}(\boldsymbol{\beta})/n$, where $l = l(\boldsymbol{\beta})$ is the usual (non-penalized) likelihood function, $\boldsymbol{i}(\boldsymbol{\beta})$ is the observed information matrix for the parameters $\boldsymbol{\beta}$, $l^* = l^*(\boldsymbol{\beta})$ is the penalized likelihood function, and $\boldsymbol{b}_{\mathbf{1}}(\boldsymbol{\beta}) = \boldsymbol{\beta}^* - \boldsymbol{\beta}^*_{\mathrm{BC}}$ for $\boldsymbol{\beta}^*_{\mathrm{BC}}$, the bias-corrected estimate. Note that $|\boldsymbol{i}(\boldsymbol{\beta})|^{1/2}$ is Jeffrey’s invariant prior for $\boldsymbol{\beta}$ which is incorporated in this score function. This utilizes no prior information about the model parameters and is useful as a method for reducing bias and avoiding infinite parameter estimates. Now consider utilizing a weakly informative prior as Gelman et al. advocate. Using this approach, the adjustment to the log likelihood is linear and has the form $(\boldsymbol{D}^{\mathbf{T}}\boldsymbol{u})^* = \partial l^*/\partial\beta = \partial l/\partial\beta - f(\sigma^2 \mid y) + \text{constant}$, where

$l = l(\boldsymbol{\beta})$ is the usual (non-penalized) log-likelihood function, $f(\sigma^2 \mid y)$ is the prior distribution, and hence the log likelihood with the penalization applied via the prior distribution is given by $l^* = l^*(\boldsymbol{\beta})$. Gelman et al. [11] proposed utilizing a three-parameter $t$ prior distribution with mean equal to 0, scale parameter equal to 2.5, and degrees of freedom equal to either 1 or 7.[12] In the 2008 manuscript, Gelman et.al. found that the $t$ distribution with $\nu = 1$ degrees of freedom (i.e. a Cauchy distribution with scale parameter 2.5) performs slightly better than the $t$ distribution with $\nu = 7$ degrees of freedom. Thus, the Cauchy prior is utilized in this study. When implementing the procedure advocated by Gelman, the explanatory variables are all centred and rescaled so that binary predictors have mean 0 and differ only by 1 with respect to the lower and upper conditions.

Both the Firth and Gelman et al. approaches provide alternative estimators for use in GLMs. Note that both estimators are functions of the model, rather than the data, as both priors are quite vague and general and, thus are applicable to almost any GLM. As the parameter estimates, either the pMLE (non-informative prior approach) or the cMLE (Cauchy weakly informative prior approach), rely on the IRLS method, the resulting parameter estimates are asymptotically multivariate normal with covariance matrix

$$\hat{\boldsymbol{V}} = \widehat{\text{cov}}(\hat{\boldsymbol{\beta}}) = (\hat{\boldsymbol{D}}^{\mathbf{T}}\hat{\boldsymbol{A}}^{\mathbf{-1}}\hat{\boldsymbol{D}})^{-1}, \tag{3}$$

where $\hat{\boldsymbol{\beta}}$ is the MLE, pMLE or Cauchy maximum-likelihood estimator (cMLE), $\hat{\boldsymbol{A}} = E(-\partial^2 l/\partial\hat{\boldsymbol{\mu}}\hat{\boldsymbol{\mu}}^{\mathrm{T}})$, and $\hat{\boldsymbol{D}} = \partial\hat{\boldsymbol{\mu}}/\partial\hat{\boldsymbol{\beta}}$ for $\hat{\mu} = g^{-1}(\boldsymbol{x}_i'\hat{\boldsymbol{\beta}})$. Thus, simultaneous intervals for the model parameters in a GLM may be constructed using these estimators and associated standard errors. However, as a consequence of utilizing the non-informative or Cauchy prior in the estimation process, the bias of the estimates are often reduced, infinite parameter estimates are avoided, and thus, these estimators may improve simultaneous estimation of functions of the parameters, particularly at smaller sample sizes.

### 2.2. *Example: a comparison of empirical coverage and length for GLM intervals*

A brief set of simulations demonstrate that the MLE intervals tend to be liberal at the small sample sizes and also have unwieldy interval lengths. Figure 1 shows the empirical error rates for estimating the slope of a GLM based on the MLE, pMLE, and cMLE across 500 simulated sets of data, with $n = 20$ and $\mu = 0.5$. The data are obtained from the simulations presented in the preceding paragraph. It is clear that the normal approximation for the MLE is not appropriate, while the normal approximations for the pMLE and cMLE are reasonable. These simulations demonstrate one reason why utilizing non-informative and weakly informative Bayesian priors should become routine when estimating GLMs, especially at smaller sample sizes. In the following, an MLE is denoted $\hat{\boldsymbol{\beta}}_{\mathrm{m}}$, a pMLE is denoted $\hat{\boldsymbol{\beta}}_{\mathrm{p}}$, and a cMLE is denoted $\hat{\boldsymbol{\beta}}_{\mathrm{c}}$. Similarly, the resulting covariance matrices for the estimates are given by $\boldsymbol{V}_{\mathrm{m}}$, $\boldsymbol{V}_{\mathrm{p}}$, and $\boldsymbol{V}_{\mathrm{c}}$, for the MLE, pMLE, and Cauchy-based covariance matrices.

### 2.3. *The estimated quantities*

Once a GLM is estimated, interest often focuses on estimating various linear combinations of the model parameters. Quantities such as the ORs or RRs are functions of the regression model parameters that may be simultaneously estimated using a matrix of coefficients.

A general expression for pointwise confidence bands on any set of linear combinations of the regression model parameters (i.e. $\boldsymbol{C}_i'\boldsymbol{\beta}$) is given by Equation (2) for $e = $ m, p, or c with $d = z_{\alpha/2}$, where $z_{\alpha/2}$ is a $z$-percentile with $\alpha/2$ area in the right tail. Note that each $\hat{\sigma}_e(\boldsymbol{C}_i)$ for $i = 1, \ldots, k$ is given by $\hat{\sigma}_e(\boldsymbol{C}_i) = \sqrt{\boldsymbol{C}_i'\hat{\boldsymbol{V}}_e\boldsymbol{C}_i}$, where $\hat{\boldsymbol{V}}_e = \boldsymbol{X}'\hat{\mathbf{W}}_e\boldsymbol{X}$ with $\hat{\mathbf{W}}_e$ the weight matrix based

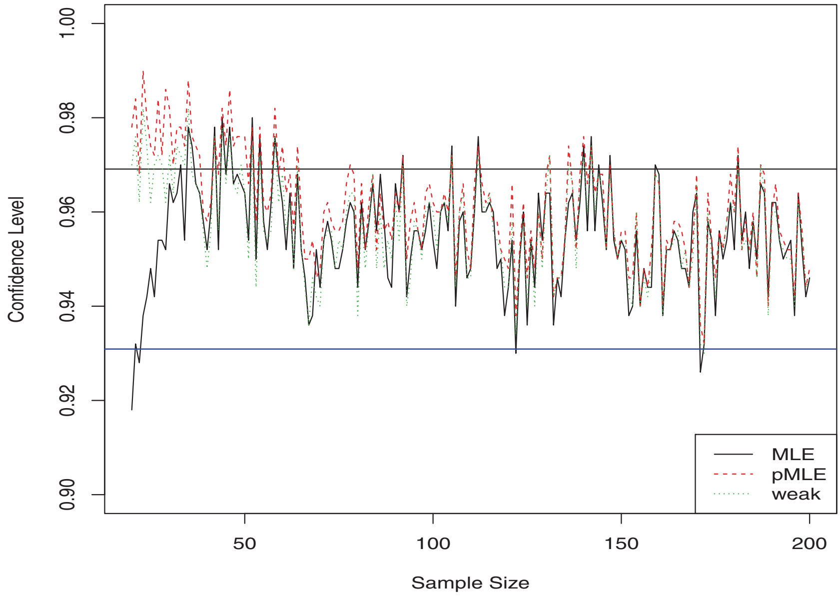


Figure 1. MLE, pMLE, and cMLE empirical type I error rates for estimating the slope parameter of a logistic regression model with $n = 20$ and $\mu = 0.5$.

on either the MLE, pMLE, or Cauchy prior estimates (denoted $e =$ m, p, or c, respectively). Let $\boldsymbol{C} = (\boldsymbol{C}_1', \ldots, \boldsymbol{C}_k')$, so that $\boldsymbol{C\beta}$ yields a set of linear combinations of the model parameters. Whenever simultaneous estimation of $\boldsymbol{C\beta}$ is desired, as in Equation (2), an adjusted critical point $d$ should be utilized in place of $z_{\alpha/2}$. We now consider various methods for determining an appropriate value for $d$ in GLM settings.

## 3. Current simultaneous estimation methods for GLM settings

When estimating quantities from a GLM there are many relevant approaches for controlling the overall error rate for a finite set of comparisons. In general, multiple comparison procedures utilize either a discrete domain approach or a continuous domain approach. For example, the Bonferroni procedure utilizes a discrete domain approach for a finite set of comparisons, but can often produce very wide intervals and is usually only suitable for a small number of comparisons. Conversely, the Scheffé procedure provides simultaneous intervals for all real values of a predictor variable which can also be used for a finite set of comparisons, thereby utilizing a continuous domain approach. Similarly, what we call the restricted-Scheffé procedure provides a continuous domain solution, suitable when only a subset of the real line is utilized and always less conservative than the traditional Scheffé procedure. This procedure, first applied to linear models by Casella and Strawderman,[6] then to logistic models by Piegorsch and Casella,[5] and then to GLMs utilizing a pMLE by Wagler and McCann,[7] can be a good solution for simultaneous inference for many points, but is often too conservative for even a large set of discrete points. Another approach makes use of resampling-based methods for multiple testing. Existing resampling-based methodologies focus on $p$-value adjustment and are often designed under closed testing assumptions.[13] These methods do not allow for inversion of the test statistics for constructing multiplicity-adjusted intervals and hence are not directly applicable to simultaneous intervals in GLMs. Westfall [14] proposed simultaneous intervals for linear regression model settings, but employs resampling of the residuals of the model. Further development of an appropriate resampling-based methodology for simultaneous intervals based on GLM parameters would seem warranted. Consequently, whenever a large number of discrete points define a fixed set of comparisons, none of these approaches (Bonferroni, Scheffé or restricted-Scheffé) usually perform well. Moreover, Bonferroni corrections can lack power for even moderate-sized sets of comparisons. Instead, a procedure is needed that controls the type I error rate without becoming too conservative.

### 3.1. *Approaches using tube-formula methodologies*

The conventional approaches for simultaneous inference outlined above are not well suited when moderate-sized finite sets of comparisons of the model parameters are of interest. In the following section, we review the tube-based SCRs applied to GLMs utilizing maximum-likelihood estimation. These tube formulas are adapted in later sections for use with the pMLE and cMLE in GLM settings. We choose to focus on the tube-based solutions for simultaneous inference in a GLM setting since a moderate-sized set of parameters are simultaneously estimated and SCRs are known to be preferable for even moderate-sized sets of inference. The performance of the SCRs is later compared to restricted-Scheffé and Bonferroni bounds. This allows for comparison with the standard continuous and discrete domain approaches, respectively. The restricted-Scheffé procedure is examined in place of the Scheffé procedure since it is always less conservative than the Scheffé procedure for restricted sets of inference.

### 3.2. *Adaptations of tube-formula SCRs for GLMs*

The simultaneous bounds developed by Sun et al. [8] utilized the MLE for parameter estimation. In later sections, we will adapt these bounds for use with the pMLE and cMLE. However, for ease in presentation, we first consider the simultaneous intervals developed by Sun et al.[8] Recall the goal is simultaneous intervals for a set of linear combinations of the model parameters, $\boldsymbol{S}_i'\boldsymbol{\beta}$ for $\boldsymbol{S}_i \in \mathcal{X}$, where the domain $\mathcal{X} \subset \mathbb{R}^k$ is a compact subset of $k$-dimensional Euclidean space. Note that a more general notation of $\boldsymbol{S}_i \in \mathcal{X}$ is utilized since the matrix need not always be a comparison matrix with whole number entries, but can be any matrix in $\mathcal{X}$.

The tube formulas proposed by Sun et al. [8] enable calculation of a value $d$ such that

$$P[\boldsymbol{S}_i'\boldsymbol{\beta} \in I(d),\ \forall \boldsymbol{S}_i \in \mathcal{X}] \geq 1 - \alpha, \quad i = 1, \ldots, k, \tag{4}$$

where $I(d)$ is a set of confidence regions for the linear combinations $\boldsymbol{S}_i'\boldsymbol{\beta}$.

In order to improve this small sample performance, Sun et al. [8] considered some modifications to the tube formula by approximating the sampling distribution of the residuals via expansions on the estimated model. Often, the vector of the observed variables ($\boldsymbol{X}_i \in \mathcal{X}$) is of interest. However, the tube-formula methodology allows estimation of any linear combination of the model parameters, such as $\boldsymbol{S}_i'\boldsymbol{\beta}$, for any vector $\boldsymbol{S}_i \in \mathcal{X}$. Consequently, $\boldsymbol{S}_i$ does not need to be observed, but just contained in $\mathcal{X}$.

Since the MLEs of the GLM parameters are known to be biased at small sizes, Sun et al. [8] consider adjustments to the critical point utilized in the SCRs. In the following, the $i$th vector of any matrix $\boldsymbol{S} = (\boldsymbol{S}_1, \ldots, \boldsymbol{S}_n)'$ is given by $\boldsymbol{S}_i = (s_{1_i}, \ldots, s_{k_i})$ and is contained in $\mathcal{X}$. This notation is utilized in describing the Gaussian process since it applies to any vector in $\mathcal{X}$, not just those defining the linear combinations of interest. All adjustments to the SCR critical point are based on expansions of the random Gaussian process $W_n(\boldsymbol{S}_i) = (g(\hat{Y}_i) - g(E(Y_i \mid \boldsymbol{S}_i)))/\hat{\sigma}_{\mathrm{m}}(\boldsymbol{S}_i)$, where $[\hat{\sigma}_{\mathrm{m}}(\boldsymbol{S}_i)]^2$ is the asymptotic variance of $g(\hat{Y}_i) = \boldsymbol{S}_i'\hat{\boldsymbol{\beta}}_{\mathrm{m}}$. The Gaussian process, $W_n(\boldsymbol{S}_i)$, converges in distribution to a Gaussian random field, $W(\boldsymbol{S}_i)$, for large $n$. Thus, the bias behaves like $|W_n(\boldsymbol{S}_i) - W(\boldsymbol{S}_i)|$. If this bias of the Gaussian field is ignored, the tube-formula critical value can be applied

Table 1. Tube-formula confidence bands.

| Interval | Formula for $g(E(Y_i \mid \boldsymbol{C}_i)) = \boldsymbol{C}_i'\boldsymbol{\beta}$ |
|---|---|
| (1) Tube SCR | $\boldsymbol{C}_i'\hat{\boldsymbol{\beta}} \pm d_{\text{TUBE}}\hat{\sigma}(\boldsymbol{C}_i)$ |
| (2) Centred SCR[a] | $\boldsymbol{C}_i'\hat{\boldsymbol{\beta}} - \hat{\kappa}_1(\boldsymbol{C}_i)\hat{\sigma}(\boldsymbol{C}_i) \pm d_{\text{TUBE}}\sqrt{\hat{\kappa}_2}\hat{\sigma}(\boldsymbol{C}_i)$ |
| (3) Corrected SCR[b] | $\boldsymbol{C}_i'\hat{\boldsymbol{\beta}} \pm d_{\text{SCRC}}\hat{\sigma}(\boldsymbol{C}_i)$ |
| (4) Corrected SCR 2[c] | $\boldsymbol{C}_i'\hat{\boldsymbol{\beta}} \pm d_{\text{SCRC2}}\hat{\sigma}(\boldsymbol{C}_i)$ |

[a]Directly utilizes the estimated centred moments $\kappa_i$ of $W_n(\boldsymbol{C}_i)$.
[b]$u$ given by $|d_{\text{SCR2}}| + q_2(\boldsymbol{C}_i) = d_{\text{TUBE}}$ and $q_2$ is a term similar to $p_2$ but based on the standardized Gaussian random field.
[c]For $d_{\text{SCRC2}} = d_{\text{TUBE}} - |\hat{r}_{\text{p}}|$ with $\hat{r}_{\text{p}} = \sup_{\boldsymbol{C}_i \in \mathcal{X}}\{p_2(\boldsymbol{C}_i, \hat{W}_n(\boldsymbol{C}_i))\} = O_p(1/n)$ is a bias estimate based on $SW_n(\boldsymbol{C}_i)$.

to the GLM with no adjustment. This is the tube SCR in Table 1 originally proposed by Sun.[15] In contrast, if this bias is utilized, three adjustments to the tube SCR are based on the inverse Edgeworth expansion of the random process given by

$$|W_n(\boldsymbol{S}_i)| = |W(\boldsymbol{S}_i)| - p_2(\boldsymbol{S}_i, W_n(\boldsymbol{S}_i)), \tag{5}$$

where the term subtracted, $p_2(\boldsymbol{S}_i, W_n(\boldsymbol{S}_i))$, is an approximation to the bias of the Gaussian process. Each of the three corrections made to the simultaneous intervals involve estimating the bias of this Gaussian process or a similar Gaussian process. Specifically, the centred SCR computes the centred moments of the process and shifts and rescales the interval accordingly (see interval (2) in Table 1). These estimated centred moments $\hat{\kappa}_i$ are based on the estimated Gaussian random field, $\hat{W}_n(\boldsymbol{S}_i)$. Intervals (3) and (4) of Table 1 approximate the bias of the standardized and unstandardized Gaussian processes and adjust the critical value based upon these approximations. These are called the corrected SCR and corrected SCR 2, respectively. The specific formulas for simultaneous bounds on a set of linear combinations of the parameters of a GLM using each of these tube-formula-based intervals is given in Table 1 and details are available in [8].

Simulations provide evidence that there is less error due to bias at small samples when either the centred or corrected versions of the intervals are utilized. Yet these intervals rely on large sample normality and correct the bias of the MLE-based parameters' post-estimation, which can introduce problems in GLM settings as demonstrated by Lin.[16] A simulation study conducted by Lin [16] investigated the coverage of the interval (4) bands of Table 1 developed by Sun et al. [8] and found that the bands had no better coverage probability than the standard hyperbolic bands where the usual Scheffé critical value is employed. However, others have successfully applied tube-formula bands to logistic GLMs for estimating multidimensional effective dose [17] and to semi-parametric logistic regression models,[18] demonstrating the utility of this methodology for GLM settings. In order to improve performance in these settings, we propose applying the tube-formula methodology to GLMs utilizing the pMLE and cMLE (in particular, the tube SCR). As the corrected and centred forms of SCRs are correcting bias introduced by the MLE, these forms are not particularly relevant when utilizing the pMLE or cMLE. We expect the pMLE and cMLE tube-formula intervals will exhibit less bias than even the corrected confidence bands proposed by Sun et al.[8] Hence, the proposed confidence bands in this manuscript should reach the desired level of confidence at smaller sample sizes than any of the aforementioned SCR methods because the pMLE and cMLE adjust for bias at any sample size, while the bias corrections employed in the SCR methods depend on asymptotic properties. Even at moderate to large sample sizes, the pMLE- or cMLE-based intervals should be competitive to the MLE-based tube-formula intervals since the MLE, pMLE, and cMLE parameter estimates have the same limiting distribution. The alternative estimators may also reduce the mean interval length when compared to the MLE-based intervals. In the following section, we apply tube-formula intervals to GLMs utilizing likelihood estimators with non-informative priors (pMLE) and weakly informative priors (cMLE) for simultaneous estimation of linear combinations of the model parameters and show that these and the MLE-based intervals provide a conservative solution for simultaneous estimation of a set of linear combinations of the parameters.

## 4. SCRs utilizing Bayesian perspective estimators

In this section, we present a framework for application of tube-formula methodology to GLMs based on Bayesian perspective estimators of the model parameters. These bounds are appropriate for making simultaneous inferences on various linear combinations of the model parameters, including quantities such as ORs or RRs.

### 4.1. *Estimating linear combinations of the model parameters*

Consider the general set of parameters and interval estimators for GLMs presented in Section 3.2. Similar bounds may be constructed utilizing pMLEs or cMLEs. Linear combinations are of the form $\boldsymbol{C}_j'\boldsymbol{\beta}$, where $\boldsymbol{C}_j$ is a $k \times 1$ vector for $j = 1, \ldots, p$.

The following theorems detail the use of the tube-formula critical values to obtain $100(1-\alpha)\%$ coverage for a fixed set of $p$ linear combinations of the parameters. This section also outlines how these solutions may be utilized in either the pMLE or cMLE context. In general, let $\mathcal{X}$ be the smallest compact subset of the domain where $\boldsymbol{C}_j \in \mathcal{X}, \forall j = 1, \ldots, p$.

THEOREM 4.1 *Under the GLM setting described in Equation* (1), *the asymptotic simultaneous coverage probability of the bands for simultaneous estimation of* $g(E(Y_i \mid \boldsymbol{C}_i)) = \boldsymbol{C}_i'\boldsymbol{\beta}$ *has a lower bound of* $1-\alpha$ *for* $d = d_{\text{TUBE}}$, $d = d_{\text{SCRC}}$ *and* $d = d_{\text{SCRC2}}$ *where these critical values are computed for* $\mathcal{X}$ *and using intervals* (1), (3), *and* (4) *of Table* 1, *respectively. For* $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{\text{p}}$, *the pMLE of* $\boldsymbol{\beta}$ *and for* $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_C$, *the cMLE of* $\boldsymbol{\beta}$, *the same result holds for* $d = d_{\text{TUBE}}$ *and interval* (1).

*Proof* Note that this result holds for any $\boldsymbol{X}_i \in \mathcal{X}$. The vectors $\boldsymbol{C}_j = (c_{j1}, \ldots, c_{jk}), j = 1, \ldots, p$, are embedded in the set $\mathcal{X}$. Thus, $\boldsymbol{C}_j \in \mathcal{X}\ \forall j$ and consequently, utilizing $d = d_{\text{TUBE}}$, $d = d_{\text{SCRC}}$, or $d_{\text{SCRC2}}$ in Equation (2) guarantees at least $100(1-\alpha)\%$ simultaneous coverage asymptotically for the intervals. Moreover, since the limiting distributions of $\hat{\boldsymbol{\beta}}_{\text{p}}$, $\hat{\boldsymbol{\beta}}_C$, and $\hat{\boldsymbol{\beta}}_{\text{m}}$ are identical, the asymptotic coverage of the bands (2) based on $\hat{\boldsymbol{\beta}}_{\text{p}}$ or $\hat{\boldsymbol{\beta}}_C$is the same as those based on $\hat{\boldsymbol{\beta}}_{\text{m}}$. ■

Another appropriate set of simultaneous intervals when estimating a linear combination of the parameters of a GLM via the pMLE or cMLE can be constructed utilizing the centred SCR (formula (2) of Table 1). The following describes how the centred SCR may be applied to pMLE- and cMLE-based estimators for this setting.

THEOREM 4.2 *Under the GLM setting described in Equation* (1), *the asymptotic simultaneous coverage probability of the band*

$$\boldsymbol{C}_j'\hat{\boldsymbol{\beta}}_{\text{m}} - \hat{\kappa}_1(\boldsymbol{C}_i)\hat{\sigma}(\boldsymbol{C}_i) \pm d_{\text{TUBE}}\hat{\sigma}(\boldsymbol{C}_i)\sqrt{\hat{\kappa}_2(\boldsymbol{C}_i)}, \tag{6}$$

*where* $\hat{\sigma}(\boldsymbol{C}_j')$ *is given by the equation in Section* (2.3), *has a lower bound of* $1-\alpha$ *with* $\hat{\kappa}_1(\boldsymbol{C}_j)$ *and* $\hat{\kappa}_2(\boldsymbol{C}_j)$ *as defined in Section* (3.2). *The same holds for* $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{\text{p}}$, *the pMLE of* $\boldsymbol{\beta}$ *with* $\hat{\kappa}_{p_1}(\boldsymbol{c}_j)$ *and*

*$\hat{\kappa}_{p_2}(\boldsymbol{C}_j)$ defined analogously for $\hat{\boldsymbol{\beta}}_{\text{p}}$, as well as for $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{\text{c}}$, the Cauchy prior estimate of $\boldsymbol{\beta}$ with $\hat{\kappa}_{c_1}(\boldsymbol{C}_j)$ and $\hat{\kappa}_{c_2}(\boldsymbol{C}_j)$ defined analogously for $\hat{\boldsymbol{\beta}}_{\text{c}}$.*

*Proof* Note that this result holds for any $\boldsymbol{X}_i \in \mathcal{X}$. The vectors $\boldsymbol{C}_j = (c_{j1}, \ldots, c_{jk}), j = 1, \ldots, p$, are embedded in the set $\mathcal{X}$. Thus, $\boldsymbol{C}_j \in \mathcal{X}\ \forall j$ and consequently the intervals in Equation (6) utilizing $\hat{\boldsymbol{\beta}}_{\text{m}}$, the usual MLE for $\boldsymbol{\beta}$, guarantee at least $100(1-\alpha)\%$ simultaneous coverage asymptotically for the intervals. Since the limiting distributions of $\hat{\boldsymbol{\beta}}_{\text{p}}$, $\hat{\boldsymbol{\beta}}_C$, and $\hat{\boldsymbol{\beta}}_{\text{m}}$ are identical, then the asymptotic coverage of the bands (2) based on $\hat{\boldsymbol{\beta}}_{\text{p}}$ or $\hat{\boldsymbol{\beta}}_C$ is the same as those based on $\hat{\boldsymbol{\beta}}_{\text{m}}$. ■

Now that methods for estimating bias-reduced linear combinations of the regression parameters for a GLM are established, they may be applied to estimate quantities such as the OR or RR. In particular, depending on the model utilized, once simultaneous bounds are computed for a set of $p$ linear combinations of the parameters of the form $\boldsymbol{C}_j'\boldsymbol{\beta}$, one may transform these results to get simultaneous bounds for a set of ORs, RRs, or attributable proportions, as these are one-to-one functions of an appropriate $\boldsymbol{C}_j'\boldsymbol{\beta}$.

## 5. Simulations

Since these methods are asymptotic, simulations provide insight into the confidence levels and comparative lengths of the intervals at various sample sizes. In the simulations, the empirical simultaneous confidence level and mean length of intervals for comparison to a control (MCC) and all-pairwise comparisons (MCA) are recorded. Recommendations are made based on these simulation results.

### 5.1. *Simulation settings*

All simulated models are of the form: $g(E(Y_i \mid \boldsymbol{X}_i)) = \boldsymbol{X}_i\boldsymbol{\beta}$, where $g$ is the link function, $\boldsymbol{\beta}$ is the $p \times 1$ vector of regression parameters, $Y_i$ is the response, and $\boldsymbol{X}_i$ is a vector of indicator variables for the $i$th subject ($i = 1, \ldots, n$). Both logistic and Poisson regression models are investigated for $p = 6$ and $p = 10$ parameters. Additionally, the probit and complementary log–log links are also examined for modelling binary responses. Reference coding is employed in the simulations so that every slope parameter is the log OR or log RR for that level of the covariate back to the reference or control. In the simulations, we consider MCC and MCA settings by selecting the appropriate comparison matrix. Assuming the reference level is the control, making comparisons to a control is equivalent to simultaneously estimating the exponents of all the slope parameters, $\exp(\beta_i)$ for $i = 1, \ldots, p$ in the model. In contrast, making MCA comparisons is equivalent to simultaneously estimating the exponent of each individual slope parameter (i.e. $\exp(\beta_i)$) and every possible difference between the model parameters, $\exp(\beta_i - \beta_j)$ for $i \neq j$, $i > j$, and $i, j = 1, \ldots, p$. In the following paragraphs, details of the simulation procedure are given.

Three scenarios are simulated for testing the proposed multiplicity corrections. The simulation settings reflect scenarios showing a small, moderate, or large change in the slopes (i.e. log OR or log RR). Each of these effect sizes are on the OR or RR scale. The parameter sets reflecting the small, moderate, and large effect sizes are (1) 50% of the $p = 6$ or $p = 10$ parameters are log(2) and the remaining are 0, (2) 50% of the $p = 6$ or $p = 10$ parameters are log(4) and the remaining are 0, and (3) 50% of the $p = 6$ or $p = 10$ parameters are log(8) and the remaining are 0, respectively. The categorical explanatory variable is based on a single binomial random variable ($T_i$), where $T_i$ is discrete uniform(0,4). To create the reference-coded indicator variables, whenever

$T_i = 0$ then $\boldsymbol{X}_i = (1, 0, 0, 0)$, if $T_i = j$ then the $j + 1$ element of $\boldsymbol{X}_i$ is 1 and all other elements are 0. This allows $T_i = 4$ to be the assumed reference level which is denoted $\boldsymbol{X}_i = (0, 0, 0, 0)$.

Once the explanatory variable matrix is generated, the response is generated in a manner appropriate for the choice of link function. When data consistent with a logit model is simulated, the response is $Y_i \sim \text{BIN}(1, \mu(\boldsymbol{X}_i))$, where each $\mu(\boldsymbol{X}_i) = E[Y_i \mid \boldsymbol{X}_i] = 1/(1 + \exp(-\boldsymbol{X}_i'\boldsymbol{\beta}))$ for $i = 1, \ldots, n$. When data consistent with a Poisson model is simulated, the response is $Y_i \sim \text{POI}(\mu(\boldsymbol{X}_i))$ with $\mu(\boldsymbol{X}_i) = E[Y_i \mid \boldsymbol{X}_i] = \exp(\boldsymbol{X}_i'\boldsymbol{\beta})$ for $i = 1, \ldots, n$. Finally, when data for probit and complementary log–log functions are simulated, the appropriate inverse distribution functions are utilized for obtaining the mean response. The sample sizes considered are $n = 50, 100, 200$, and 300 for all sets of comparisons. All simulations assume a $\alpha = 0.05$ family-wise error rate and are performed in *R* [19] using the *glm* function for model estimation. In all, 1000 data sets were simulated for each parameter set, model, and sample size combination.

Recall that MLEs are prone to convergence problems when the data are linearly separable. Since linear separability occurs more frequently at smaller sample sizes, estimation warnings occurred when utilizing the MLE in the simulations. In this event, a warning that the algorithm fails to converge is returned by the software and is subsequently recorded in the simulations. Table 2 summarizes the number of warnings for each setting. In addition to the warnings reported by the software during parameter estimation, the MLE may yield very large parameter estimates without returning a warning message. Whenever the mean length of the interval exceeded 20, the case was recorded as an estimation error but still retained in the simulation records. In the simulation results presented in the next section, empirical confidence levels and interval lengths are given for the tube SCR when utilizing the MLE, pMLE, and cMLE. Though the theorems in Section 4.1 allow other intervals, such as the centred pMLE- or cMLE-based intervals and all four versions of the MLE-based SCRs, simulations suggest that the corrected and centred versions of the SCR proposed by Sun et al. [8] did not yield an improvement over the tube-based formulas in this setting. Hence, just the tube formulas applied to the MLE, pMLE, and cMLE are presented in the following sections. In order to provide a relevant comparison, MLE-based Bonferroni-adjusted intervals are also presented. This is an appropriate comparison since the Bonferroni procedure is widely utilized and is known to be a conservative option. The simulations used the Bonferroni corrections for all estimators (MLE, pMLE, and cMLE) considered. The empirical error is only reported for the MLE-Bonferroni results because the pMLE-Bonferroni and cMLE-Bonferroni results were always within 0.004 of the MLE-Bonferroni empirical error rates across all of the simulation settings.

Table 2. Counts of MLE non-convergence errors and separability warnings for $p$ =6 and 10 parameters based on 500 simulated data sets.

| Model | Parameter set | $n$ | Non-convergent | Warning ($p = 6$) | Warning ($p = 10$) |
|---|---|---|---|---|---|
| Logistic | Large effect | 50 | 1 | 216 | 130 |
| Logistic | Large effect | 100 | 0 | 14 | 6 |
| Poisson | Large effect | 50 | 4 | 35 | 142 |
| Poisson | Large effect | 100 | 0 | 1 | 6 |
| Logistic | Moderate effect | 50 | 4 | 155 | 127 |
| Logistic | Moderate effect | 100 | 0 | 4 | 1 |
| Poisson | Moderate effect | 50 | 2 | 35 | 138 |
| Poisson | Moderate effect | 100 | 0 | 0 | 6 |
| Logistic | Small effect | 50 | 4 | 281 | 130 |
| Logistic | Small effect | 100 | 0 | 13 | 8 |
| Poisson | Small effect | 50 | 1 | 25 | 142 |
| Poisson | Small effect | 100 | 0 | 0 | 12 |

Note: There were no non-convergence errors when $p = 6$.

### 5.2. *Simulation results*

In the following, the MCC and MCA simulation results are analysed. Figures 2 and 3 display the simulation results for the MCC and MCA settings with $p = 6$ parameters and Figures 4 and 5 show these results for 10 parameters where the empirical errors are plotted as a function of the ratio of sample size to number of parameters ($n/p$) for various interval estimators. Note that whenever the sample size to number of parameters is small ($n/p < 10$), the simultaneous intervals are all uniformly conservative. This is true for all critical points and estimation methods evaluated and is due to the larger than necessary standard errors associated with the parameter estimates when there is a relatively small number of observations. For all MCC cases, the MLE has superior performance for most cases whenever $n/p \geq 30$. In contrast, the cMLE performs best or close to best for $n/p \leq 30$. In general, the Bonferroni-based intervals are slightly more conservative for MCC. The difference between the tube-based intervals and Bonferroni is not large for MCC due to the relatively few number of comparisons being made. For MCA cases, the Bonferroni is extremely conservative in all scenarios considered and, in general, the methods perform similarly in all Poisson MCA scenarios. The results are somewhat different for $p = 6$ and $p = 10$ when considering just the logit model. For the logit MCA cases, there can be a slight improvement when utilizing the cMLE when $p = 6$. This is most evident at small sample sizes, and in general, the competing methods perform similarly for all scenarios at the larger sample sizes. When $p = 10$, the cMLE appears best for all sample sizes when the effect size is large, and best for $n/p \leq 10$ for small and medium effects. Whenever $n/p > 10$ the MLE performs slightly better for small and medium effects. Binomial GLMs were also evaluated using probit and complementary log–log link functions. These links yielded results very similar to the logit link results and, therefore, are omitted from the plots in this section and the following. However, it should be noted that the

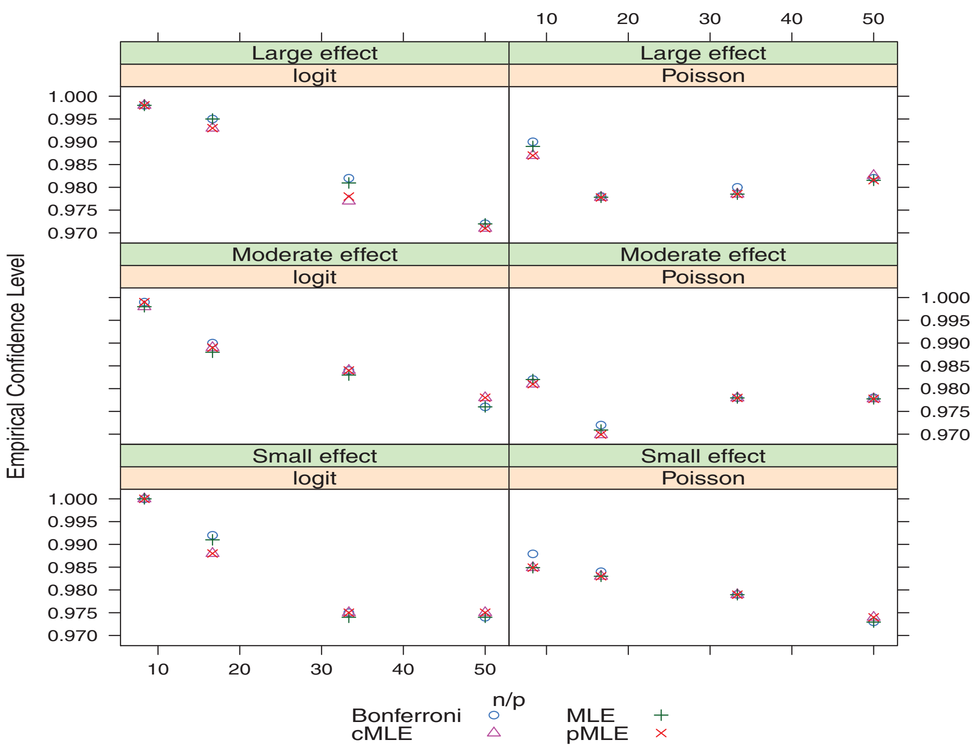


Figure 2. Empirical confidence levels for 95% MCC simultaneous intervals with $p = 6$ based on 1000 simulations.

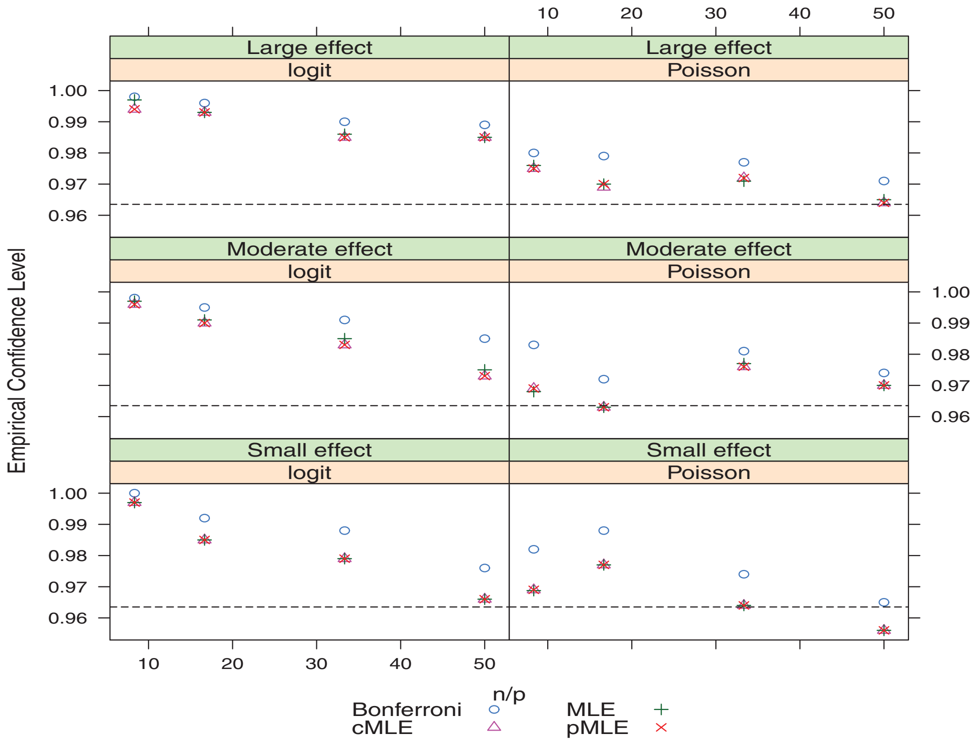


Figure 3. Empirical confidence levels for 95% MCA simultaneous intervals with $p = 6$ based on 1000 simulations (dashed line indicates bounds for obtaining 95% confidence).

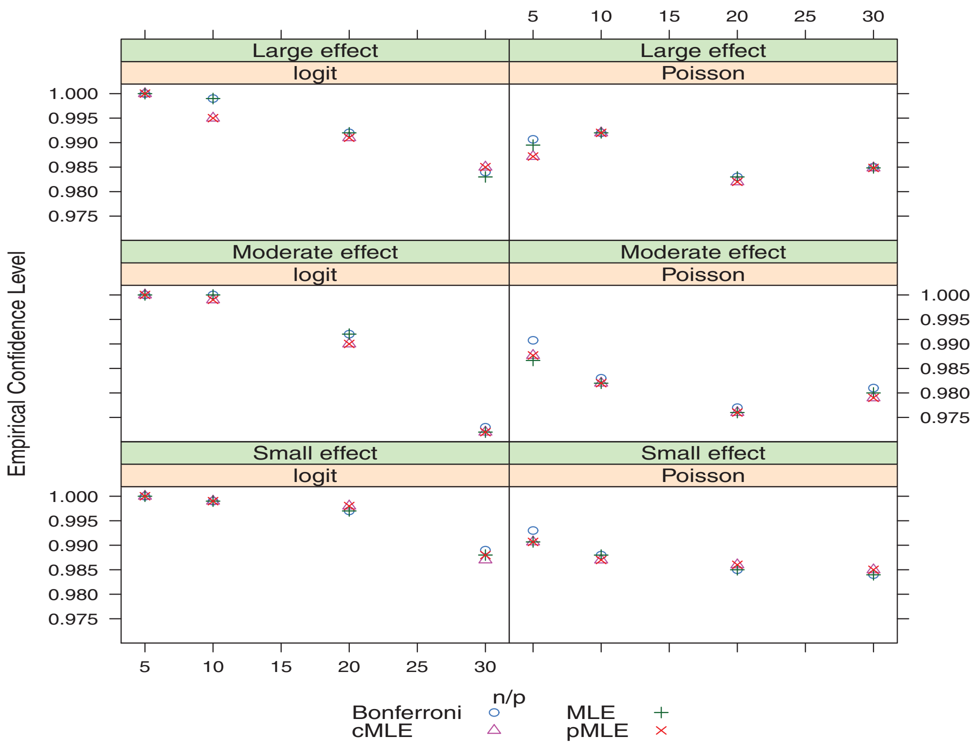


Figure 4. Empirical confidence levels for 95% MCC simultaneous intervals with $p = 10$ based on 1000 simulations.

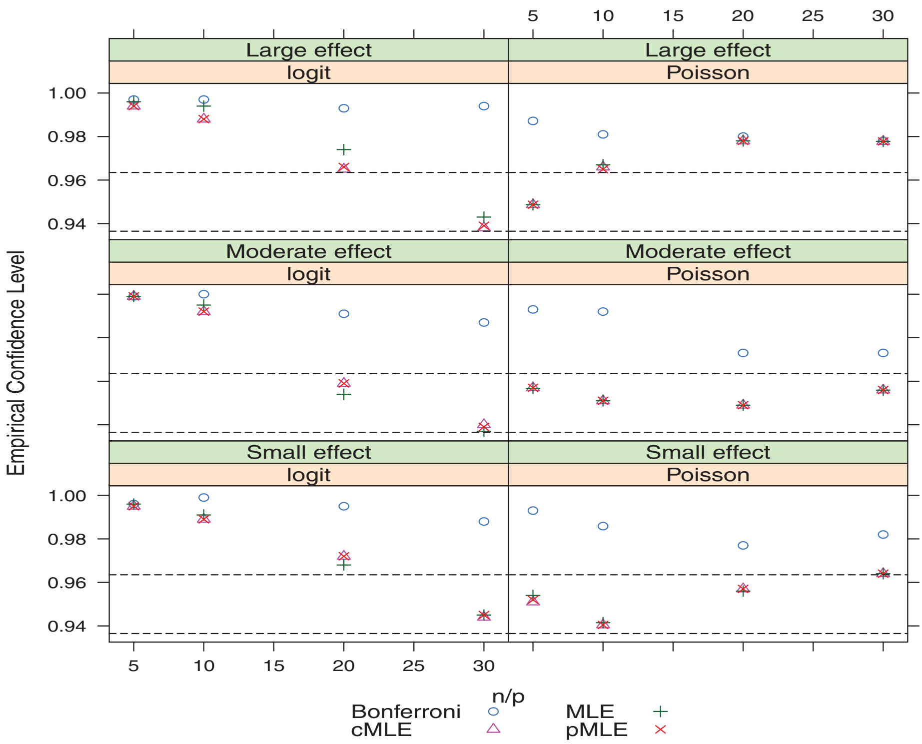


Figure 5. Empirical confidence levels for 95% MCA simultaneous intervals with $p = 10$ based on 1000 simulations (dashed line indicates bounds for obtaining 95% confidence).

probit link was slightly less conservative than the logit link while still maintaining the desired error rate.

### 5.2.1. *Comparisons of interval length*

In some of the previous results, the pMLE-, cMLE-, and MLE-based intervals tend to perform similarly when considering just the empirical confidence level. In such cases, it is appropriate to consider interval length when selecting a method for simultaneous inference. Figures 6 and 7 present violin density plots of the lengths for the simultaneous intervals of $\boldsymbol{C}_i'\boldsymbol{\beta}$ for logistic and Poisson regression models with $p = 6$, assuming the second parameter set ($\boldsymbol{\beta} = (0, 0, 0, \log(4), \log(4), \log(4))$) and sample sizes $n = 50, 100, 200$, and 300. The $\boldsymbol{C}_i'\boldsymbol{\beta}$ are appropriate for MCA and we utilize tube formulas for MLE-, pMLE-, and cMLE-based intervals. For display, the lengths are compared only for the cases where the MLE-based interval lengths were 20 units or less. This provides a fair comparison since the MLE would not be used when exhibiting characteristics of separability. (See Table 2 for a summary of the cases with lengths greater than 20.) In addition to presenting the lengths, the mean value for the tube-formula critical value is also given for each model and sample size simulated.

The simulation results indicate that at any sample size, the MLE-based intervals were longest, followed by the pMLE, and then the cMLE-based intervals. At small and moderate sample sizes such as $n = 50$ and $n = 100$, the obtained MLEs and associated standard errors can be very large for the logit and Poisson models, resulting in intervals that are wide. In contrast, the cMLE- and pMLE-based simultaneous intervals maintain more reasonable lengths and standard errors at the same sample sizes. Once the sample size is larger ($n = 200$), the MLEs less frequently yield

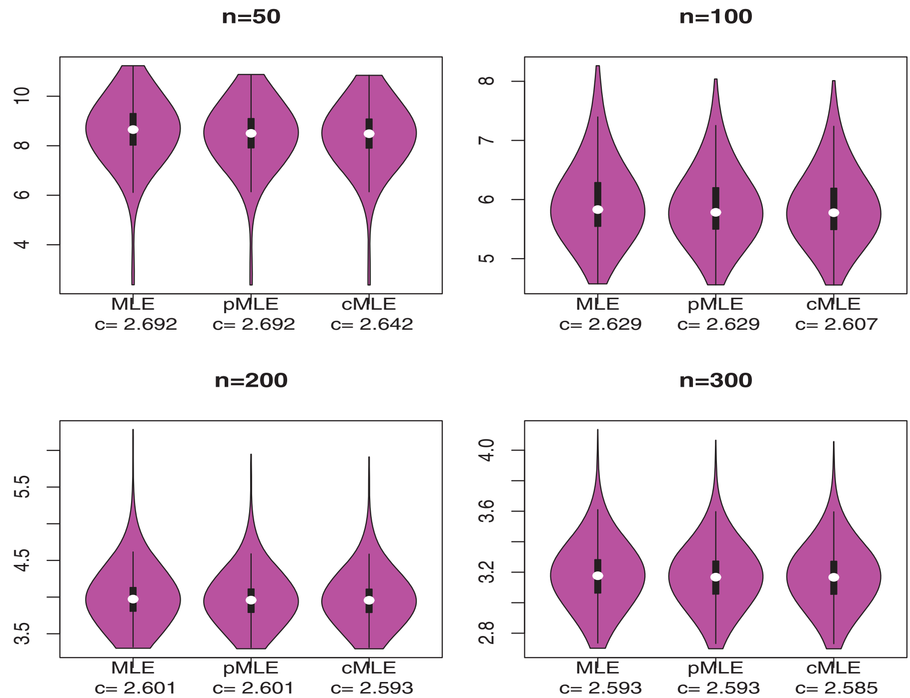


Figure 6. A comparison of simultaneous interval length for a logistic model using parameter set 2 (moderate effect) and $p = 6$ (from left to right $n/p = 50$, 100, 200, and 300).

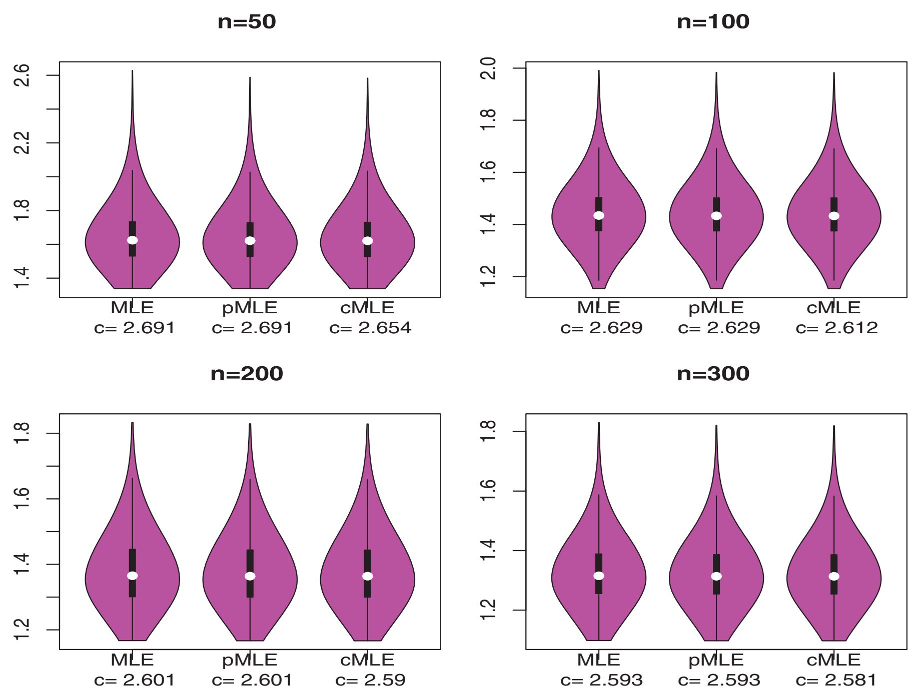


Figure 7. A comparison of simultaneous interval length for a Poisson model using parameter set 2 (moderate effect) and $p = 6$ (from left to right $n/p = 50$, 100, 200 and 300).

substantially larger standard errors and the MLE-based interval lengths are much closer to the cMLE- or pMLE-based interval lengths for both logit and Poisson models. This provides further evidence that an alternative estimator should be utilized in place of the MLE for GLMs whenever the sample size is small and that the choice of an alternative estimator does not negatively affect the calculation of the tube-based critical value. In fact, the shorter length is also partly due to the decreased size of the tube-based critical value when utilizing the pMLE or cMLE estimators (Figures 6 and 7). As the simulations suggest, for small to medium sample sizes, the cMLE- and pMLE-based intervals yield empirical confidence levels that are close to 95% and the pMLE and cMLE interval lengths are shorter than the MLE-based intervals even when excluding the cases where the MLE is inappropriately long (i.e. length $> 20$). In consideration of these and the previous empirical error rates, the cMLE appears preferable at smaller sample sizes ($n \leq 200$) as it maintains the desired level of confidence with the greatest level of precision. Once $n > 200$ the MLE performs well and is the most convenient estimator to use in GLM settings, excluding cases exhibiting separability.

#### 5.2.2. *Recommendations*

Due to the estimation problems with the MLE and occasional correspondingly large standard errors, it is prudent to use the cMLE with the naive tube formula whenever $n \leq 200$. Although the cMLE is recommended for these small to moderate sample sizes, for larger samples the regular tube-based intervals utilizing the MLE is a reasonable solution though can result in wider intervals. When only a small number of comparisons are desired and $n \leq 200$, using the cMLE along with the tube-formula critical value is optimal, but substituting the Bonferroni critical value would be acceptable. However, in any setting, the tube formulas provide a significant increase in power when the number of comparisons is larger (more than 5 comparisons).

## 6. Applications

### 6.1. *Logistic regression: childhood asthma*

We illustrate application of these penalized simultaneous bounds using the 2009 NHIS [1] data. The explanatory variables in the model are all indicator variables specifying region (NE, MW, S, W) and whether the child is diagnosed with hayfever allergies. It is reasonable to make comparisons across the regions for children with and without diagnosed allergies. The model is estimated using MLEs, pMLEs, and cMLEs and both models have the form $\text{logit}(\mu_i) = \alpha + \beta_1 x_{\text{MW}} + \beta_2 x_{\text{S}} + \beta_3 x_{\text{W}} + \beta_4 x_{\text{noallergy}}$, where each $x_i$ is an indicator variable for the specified condition. Table 3 displays the estimated confidence intervals for the linear combinations of the parameters that are of interest. Values where the OR is different than 1 are marked with an asterisk.

The MLE-based intervals significantly differ from both the pMLE- and cMLE-based intervals due to the larger standard errors associated with MLEs. For these data, the MLE-based intervals yielded reasonable values and do not exhibit signs of separation. The factors identified as significantly different are all the same across the competing intervals except when the northeast (NE) and south (S) are compared for children with no diagnosed hayfever allergies and when the south (S) and west (W) are compared for children with diagnosed allergies. For the first case, the cMLE-based OR interval estimate does not contain 1, while the MLE- and pMLE-based OR intervals do contain 1, and for the second case, the MLE-based interval is the only interval that does not contain 1. Recall that the lengths of the simultaneous intervals are reduced when utilizing the cMLE rather than the pMLE or MLE due to a smaller critical value on average and reduced

Table 3. Logistic regression for predicting childhood asthma.

| Region | Hayfever | log OR | Tube MLE OR | Tube pMLE OR | Tube cMLE OR |
|---|---|---|---|---|---|
| NE vs. MW | No | $\beta_1$ | (7.90, 9.11)* | (6.09, 6.98)* | (4.90, 5.50)* |
| NE vs. S | No | $\beta_2$ | (0.96, 1.08) | (0.91, 1.02) | (0.75, 0.82)* |
| NE vs. W | No | $\beta_3$ | (2.16, 2.44)* | (1.95, 2.19)* | (1.59, 1.76)* |
| MW vs. S | No | $\beta_1 - \beta_2$ | (7.85, 8.84)* | (6.38, 7.14)* | (6.27, 6.99)* |
| MW vs. W | No | $\beta_2 - \beta_3$ | (0.42, 0.46)* | (0.45, 0.49)* | (0.45, 0.49)* |
| S vs. W | No | $\beta_1 - \beta_3$ | (3.47, 3.93)* | (2.97, 3.35)* | (2.93, 3.29)* |
| NE vs. MW | Yes | $\beta_1 + \beta_4$ | (1.81, 2.13)* | (1.76, 2.05)* | (1.55, 1.79)* |
| NE vs. S | Yes | $\beta_2 + \beta_4$ | (0.22, 0.25)* | (0.26, 0.30)* | (0.24, 0.27)* |
| NE vs. W | Yes | $\beta_3 + \beta_4$ | (0.50, 0.57)* | (0.56, 0.65)* | (0.50, 0.57)* |
| MW vs. S | Yes | $\beta_1 - \beta_2 + \beta_4$ | (1.78, 2.10)* | (1.82, 2.13)* | (1.97, 2.29)* |
| MW vs. W | Yes | $\beta_2 - \beta_3 + \beta_4$ | (0.10, 0.11)* | (0.13, 0.15)* | (0.14, 0.16)* |
| S vs. W | Yes | $\beta_1 - \beta_3 + \beta_4$ | (0.79, 0.93)* | (0.85, 1.00) | (0.92, 1.08) |

*Interval is statistically significant.



standard errors. Recall too that the empirical error rates are very similar in almost all cases. For this example, the tube critical values were $c_{\text{MLE}} = 2.383$, $c_{\text{pMLE}} = 2.3683$, and $c_{\text{cMLE}} = 2.384$. The usual Scheffé critical value is 3.49 and the Bonferroni critical value is 2.865. Using either would have yielded far more conservative intervals. Even with a slightly larger tube critical value, the cMLE estimators yielded the most precise simultaneous interval estimators due to the associated decrease in the standard errors for $\boldsymbol{C}_i'\boldsymbol{\beta}$. Note that, for this scenario, the simulations support using the cMLE-based tube intervals and for these data it does return more significant associations than the MLE-based intervals.

### 6.2. *Poisson regression: predicting days missed*

Using the same 2009 NHIS [1] data, a Poisson regression model is employed to predict days missed due to illness. The population of interest is all children ever diagnosed with asthma and less than seven years of age. The Poisson regression model uses covariates region (as coded in the previous example) and still have asthma (whether or not the child still has asthma among children less than seven years of age who have ever been diagnosed with asthma). Allowing for interaction terms between region and the indicator variable for still having asthma, the Poisson regression model is of the form $\log(\mu_i) = \beta_1 x_{\text{asthma}} + \beta_2 x_{\text{none}} + \beta_3 x_{\text{MW}} + \beta_4 x_{\text{S}} + \beta_5 x_{\text{W}} + \beta_6 x_{\text{asthma}} * x_{\text{MW}} + \beta_7 x_{\text{asthma}} * x_{\text{S}} + \beta_8 x_{\text{asthma}} * x_{\text{W}}$, where each $x_i$ is an indicator variable for still having asthma or region. Table 4 presents the resulting model-based RRs for assessing which combinations of region and the still have asthma variable are useful for predicting the number of school days missed due to illness for the $n = 57$ observations.

Table 4. Poisson regression for predicting days missed.

| Region | Still have asthma | log RR | Tube MLE | Tube pMLE | Tube cMLE |
|---|---|---|---|---|---|
| NE | Yes | $\beta_1$ | (7.08, 7.41)* | (7.09, 7.40)* | (7.09, 7.40)* |
| NE | No | $\beta_2$ | (8.81, 9.25)* | (8.82, 9.18)* | (8.82, 9.17)* |
| MW | Yes | $\beta_1 + \beta_3$ | (0.72, 1.11) | (0.73, 1.11) | (0.73, 1.11) |
| MW | No | $\beta_2 + \beta_3 + \beta_6$ | (1.68, 2.38)* | (1.70, 2.37)* | (1.70, 2.37)* |
| S | Yes | $\beta_1 + \beta_4$ | (6.08, 6.28)* | (6.08, 6.27)* | (6.08, 6.27)* |
| S | No | $\beta_2 + \beta_4 + \beta_7$ | (0.82, 1.22) | (0.83, 1.22) | (0.83, 1.22) |
| W | Yes | $\beta_1 + \beta_5$ | (8.48, 8.82)* | (8.49, 8.81)* | (8.49, 8.81)* |
| W | No | $\beta_2 + \beta_5 + \beta_8$ | (7.77, 8.66)* | (7.79, 8.63)* | (7.79, 8.63)* |

*Interval is statistically significant.

While the research conclusions would not substantively change with regard to which interval method is utilized, we note that the interval length is shortest for the pMLE- or cMLE-based intervals. Moreover, the Bonferroni value for this set of inferences would be 2.734, while the Tube MLE is 2.621, Tube pMLE is 2.490, and Tube cMLE is 2.490. Thus, utilizing the cMLE or pMLE estimator allows for more precise interval estimates.

## 7. Conclusions

Epidemiological and medical research routinely employs GLMs where interest is often focused on quantities estimated from the model such as ORs or RRs. When simultaneous inference of these quantities is warranted, conventional methods for controlling the family-wise error rate, such as Bonferroni or Scheffé, are often not ideal. Instead, a method that is more appropriate for simultaneous inference on model parameters estimated with a moderate-sized set of discrete points is desirable. Additionally, at small to moderate sample sizes the MLE is known to produce biased parameter estimates and often does not even yield 'reasonable' parameter estimates, especially when the explanatory variables are categorical. Thus, a method for simultaneous inference on multiple ORs, RRs, or similar quantities that takes into account the estimation problems prevalent when there are only categorical predictors in the model is desirable.

In this manuscript, we present a method for simultaneous estimation of quantities estimated via linear combinations of the regression parameters that is particularly suitable at small to moderate sample sizes. These bounds utilize either the pMLE or cMLE as an alternative to the MLE and also employ tube formulas for calculating the simultaneous critical value. The pMLE and cMLE are utilized because these estimators can provide finite parameter estimates that are less biased than the MLE-based estimates. Additionally, the proposed intervals utilize a method for simultaneous inference based on tube-formula approximations. The tube-based SCRs, applied to MLE-based GLMs by Sun et al.,[8] are typically utilized for continuous domains, but may also be applied in a setting with categorical predictor variables. When a moderate to large set of discrete points defines the set of parameters to be simultaneously estimated, the tube-formula bounds are often a better solution than any of the competitors and, when employed with the either the cMLE or pMLE, provide a powerful solution even for small to moderate sample sizes. Simulations suggest that the proposed bounds reach the desired level of confidence and are less conservative than other methods for simultaneously estimating linear functions of the model parameters. The procedure is also shown to reduce the length of the simultaneous bounds on these sets of parameters at smaller sample sizes. Consequently, when estimating linear functions of GLM parameters, the cMLE with the naive tube-formula critical value is recommended when $n \leq 200$, while the MLE-based tube intervals are acceptable for larger sample sizes. In practice, these intervals may be computed in *R* using the *locfit* package.[20] Additionally, pMLE and cMLE routines are also available in *R* in the *brglm* package and using the Gelman function *bayesglm*.

## References


[1] Division of Health, National Health Interview Survey (NHIS) public use data release NHIS survey description. Hyattsville, MD: Division of Health Interview Statistics, National Center for Health Statistics; 2009.

[2] Hsu J. Multiple comparisons: theory and methods. 1st ed. Boca Raton, FL: CRC; 1996.

[3] Worsley K. An improved Bonferroni inequality and applications. Biometrika. 1982;69:297–302.

[4] Hunter D. An upper bound for the probability of a union. J Appl Probab. 1976;13:597–603.

[5] Piegorsch D, Casella G. Confidence bands for logistic regression with restricted predictor variables. Biometrics. 1988;4:739–750.

[6] Casella G, Strawderman W. Confidence bands for linear regression with restricted predictor variables. J Am Statist Assoc. 1980;75:862–868.

[7] Wagler A, McCann M. Bias-reduced simultaneous confidence bands on generalized linear models with restricted predictor variables. J Statist Theory Practice 2012;6(2):286–302.
[8] Sun J, Loader S, McCormick D. Confidence bands in generalized linear models. Ann Stat. 2000;28(2):429–460.
[9] McCullagh P, Nelder J. Generalized linear models. 2nd ed. London: Chapman and Hall; 1989.
[10] Firth D. Bias reduction of maximum likelihood estimates. Biometrika. 1993;80(1):27–38.
[11] Gelman A, Jakulin A, Grazia Pittau M, Su Y. A weakly informative default prior distribution for logistic and other regression models. Ann Appl Stat. 2008;2(4):1360–1383.
[12] Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian data analysis. London: Chapman and Hall; 1995.
[13] Westfall P, Troendle J. Multiple testing with minimal assumptions. Biom J. 2008;50:745–755.
[14] Westfall P. On using the bootstrap for multiple comparisons. J Biopharm Stat. 2011;21(6):1187–1205.
[15] Sun J. Tail probabilities of the maxima of Gaussian random fields. Ann Probab. 1993;21:34–71.
[16] Lin S. Simultaneous confidence bands for linear and logistic regression models [Ph.D. thesis]. Southampton: University of Southampton; 2008.
[17] Li J, Nordheim EV, Zhang C, Lehner C. Estimation and confidence regions for multi-dimensional effective dose. Biom J. 2008;50(1):110–122.
[18] Li J, Zhang C, Doksum K, Nordheim E. Simultaneous confidence intervals for semi-parametric logistic regression and confidence regions for the multi-dimensional effective dose. Statist Sinica. 2010;20(2):637–659.
[19] R.D.C. Team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2010. ISBN 3-900051-07-0. http://www.R-project.org/
[20] Loader C. Locfit: local regression, likelihood and density estimation. R package version 1.5-9.1; 2013. http://CRAN.R-project.org/package=locfit